Extended Project Qualification

To what extent does AI assist in identifying exoplanets using the transit photometry method?

Name: **Roman Kirkorov**

Candidate Number: **1083**

School: **Cardiff Sixth Form College Cambridge**

Table of contents

Introduction

Astronomy, the ancient science of studying celestial objects and phenomena, has enthralled human curiosity for thousands and millions of years. At its core, it explores the vastness of the cosmos, encompassing everything from the smallest particles to the largest galaxies. One of the most compelling quests within astronomy is the search for exoplanets – planets orbiting stars beyond our solar system. This mission fuels our imagination about whether there is life beyond Earth, and allows us to understand the formation and evolution of planetary systems. According to Batler (2022), astrophysics is “a branch of space science that applies the laws of physics and chemistry to seek to understand the universe and our place in it”. Specifically, it studies the evolution of the universe, the behaviour of stars, galaxies, and other cosmic structures (Balter, 2022). From the nuclear fusion reactions powering stars to the gravitational interactions shaping galaxies, astrophysics provides crucial insights into the fundamental laws of nature operating on cosmic scales. So far, humanity only understands 5%, which is ordinary matter, of the universe because the other 95% (dark energy and dark matter) are completely enigmatic (Barberio, 2020). The study of exoplanets sits at the intersection of astronomy and astrophysics, trying to expand the boundaries of how humans understand planetary systems. Detecting and characterising these distant worlds requires advanced observational techniques and computational models. In recent years, artificial intelligence (AI) has emerged as a powerful tool in this endeavour, revolutionising the way astronomers search for and analyse exoplanet data. In this research, I intend to evaluate the role AI plays in identifying planets outside the Solar System using the transit photometry method. I want to study this topic specifically for several reasons. Initially, my father’s background in aerospace engineering ignited my childhood fascination with space exploration, fostering a deep interest in observing the night sky. Secondly, delving into machine learning is particularly pertinent today as it is an emerging field with ample opportunities for exploration.

Lastly, my aspiration to pursue Computer Science at university aligns well with my interest in AI and its potential for automating searches, and wish to further explore its applications, especially in astronomy. After a brief introduction to astronomy and exoplanets, this essay will explore the details of the transit photometry method's principle of operation and its strengths and weaknesses. Next, the essay will cover the machine learning basics, before analysing and evaluating various studies on this particular topic, acquiring an understanding of how AI is beneficial for this purpose. The research question will ultimately be answered after thorough research on the topic.

Astronomy and the physics concept

In this phase of the project, I will delve into the physics pertaining to my subject matter. This will involve an introduction to exoplanets and an exploration of the mechanics associated with utilising the transit photometry method.

Exoplanets and their identification

Initially, it is essential to establish a precise definition of the term ‘planet’ to provide a clear framework for delineating the characteristics of the exoplanet under discussion. The International Astronomical Union (2018) defines a planet as “A celestial body that (a) is in orbit around the Sun, (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, and (c) has cleared the neighbourhood around its orbit”. Therefore, an exoplanet is as a celestial body that revolves around a star other than our Sun (Lissauer, 2024). The significance of exoplanets and their search lies in their potential to harbour life beyond Earth. Moreover, observing the formation of other star systems provides valuable insights into the formation of our own system, our history, and the general functioning of the universe (The Planetary Society, n.d.). The discovery of a new planet evokes a sense of wonder, suggesting the possibility of other beings contemplating the workings of the cosmos and pondering their existence.

The first exoplanets were jointly identified in 1992 by astronomers Aleksander Wolszczan and Dale Frail (Wenz, 2023). Since then, the search for exoplanets has gained momentum. With the introduction of NASA’s Transiting Exoplanet Survey Satellite (TESS) and the Webb Space Telescope, the rate of discovery has increased (Winn, 2023). TESS employs four identical, highly optimised, red-sensitive (able to detect infrared radiation), wide-field cameras to observe the sky (MIT and the Centre for Astrophysics, 2018), while the Webb Space Telescope, launched in 2021, detects infrared radiation from stars, nebulae, galaxies, planets, and exoplanets using sensitive sensors (Webb Space Telescope, 2024). Consequently, advancements in observational technologies and space object detection, such as advanced telescopes, have expedited and improved the identification of exoplanets, thereby enhancing our understanding of the characteristics of extraterrestrial life.

Transit photometry method and its principle of operation

There are five methods for detecting exoplanets: radial velocity, direct imaging, microlensing, astrometry, and transit photometry (European Space Agency, 2019). However, the transit photometry method is currently the most successful, effective, and sensitive (The Planetary Society, 2020). The reason for focusing on researching this method specifically lies in its popularity and the availability of data. At the time of writing, NASA has used this method to discover over 4150 planets, whereas a total of fewer than 1400 exoplanets were identified through the other four methods combined (NASA Exoplanet Exploration, 2024), indicating its superior performance. It possesses several advantages, which will be discussed after understanding its working mechanism. For now, the other four methods will be briefly introduced for a better general understanding. Radial velocity involves observing stars' wobbling caused by orbiting planets, which alters the colour of light seen by astronomers (NASA Exoplanet Exploration, 2024). Direct imaging captures photographs of exoplanets by suppressing the bright light from the stars they orbit, while gravitational microlensing and astrometry are used by observing changes related to gravity and distance, respectively (NASA Exoplanet Exploration, 2024).

In the term 'transit photometry,' the word 'transit' denotes the passage of a planet in front of the star, while 'photometry' refers to the measurement of incoming light. The operational principle of the transit photometry method is straightforward and intuitive. It relies on the concept of a star and an orbiting exoplanet, and works similarly to noticing a slight dimming of a light bulb when a small object moves in front of it. When the exoplanet transits or passes in front of the star, the brightness of the star slightly decreases as less light reaches the observer (NASA Exoplanet Exploration, 2024). The duration of the transit, the interval between transits, and the size of the planet provide valuable information about the planetary system (NASA Exoplanet Exploration, 2024). Figure 1 illustrates the transit cycle (British Astronomical Association, 2022), which will now be closely analysed.

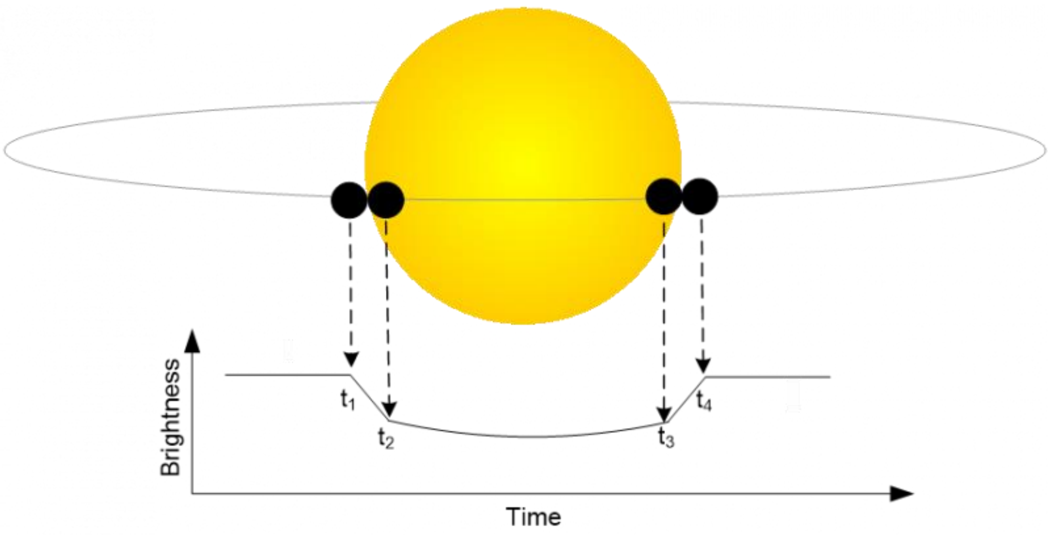

Figure 1: An exoplanet transit (Source: British Astronomical Association, 2022)

Initially, the brightness of the star remains at its maximum and constant until the planet becomes visible at time t_1, known as external ingress. Between t_1 and t_2, the planet enters the observer's line of sight, causing a gradual decrease in luminosity. At time t_2, referred to as internal ingress, the planet is fully within the star's silhouette from the observer's perspective. Throughout the period between times t_2 and t_3, the brightness remains relatively constant; however, as the planet obstructs the star's core more significantly, the signal dims slightly. This is because the nuclear fusion processes generating light occur predominantly in the star's core (American Museum of Natural History, 2024). Therefore, when the core is obstructed, the primary source of light is impeded, resulting in less light reaching the observer when the core is not obstructed.

Subsequently, following t_3 (internal egress) and before t_4 (external egress), the planet moves out of the star’s vicinity, leading to a gradual increase in overall luminosity. Finally, between t_4 and t_1 of the next cycle, the planet orbits around the star, a movement imperceptible to the observer. By accurately plotting the luminosity-time graph and minimising interference, it becomes possible to determine the exoplanet’s orbital period or year.

Advantages of the transit photometry method

As indicated, the transit photometry method holds advantages over other techniques. Primarily, it provides precise estimations of the sizes of discovered planets, which would otherwise be challenging to ascertain (Williams, 2017). Additionally, the transit method can be combined with the radial velocity method for determining the planet’s density, offering valuable insights for future studies on the exoplanet’s composition (Williams, 2017). Moreover, it proves particularly advantageous for space-based observatories capable of continuous observation of stars over extended periods, spanning weeks or even months (The Planetary Society, 2020). For instance, the TESS spacecraft maintains focus on its targets for durations ranging from a minimum of 27 days (almost one month) to a maximum of 351 days (almost one year) (MIT and the Centre for Astrophysics, 2018). To further investigate an exoplanet, spectroscopy techniques can be employed to create an absorption spectrum, revealing the composition of the planet’s atmosphere by examining light curves across various wavelengths (The Planetary Society, 2020). This spectroscopic approach is particularly beneficial for assessing the potential habitability of exoplanets, with a specific focus on detecting indicators of the greenhouse effect (Seager, 2015). Finally, transit photometry enables the identification of exoplanets located several thousand light-years away from their host stars, a capability not shared by other detection methods (Gotame, 2020). Thus, it is evident that this method offers a range of advantages, rendering it a reasonably favourable choice.

Problems emerging through the process using this method

However, the transit photometry method also presents several challenges throughout the exoplanet identification process. The primary and most significant drawback is the requirement for any distant planet to pass through the line formed by the star and Earth while staying on its orbit; not every planet that is orbiting another star performs such transits as observed from Earth (Williams, 2017). Unfortunately, only around a tenth of distant planets has orbits nearly perfectly aligned with the observer, which is essential for a successful transit (Williams, 2017), leaving a multitude of planets undetectable by photometry. Secondly, while a planet's orbital period may span months or years, the duration of a transit typically lasts only a few hours or days (The Planetary Society, 2020). Furthermore, when a spacecraft transmits collected data to Earth, it ceases observation of the surroundings to transmit the data, potentially missing rare transits occurring during this period. Consequently, short-period exoplanets are easier to identify. Thirdly, there is a relatively high incidence of false positives, where objects mistakenly identified as exoplanets turn out not to be. Research conducted by Santerne et al. (2012) suggests that false positives for transits detected by the Kepler mission, which relied on the transit technique, may constitute up to 40%. This issue necessitates astronomers to observe multiple regular transits to confirm the presence of a planet (The Planetary Society, 2020). Therefore, exoplanets identified using the transit technique typically require further confirmation, often through methods such as orbital brightness modulation or radial velocity (Gotame, 2020). Lastly, factors such as mixed signals or interference from other celestial objects such as asteroids can significantly increase noise in the luminosity-time graph, making it more challenging to interpret and analyse. For example, if a planetary system comprises three planets with varying orbital lengths and periods, it could significantly impact the brightness-time graph (NASA Exoplanet Exploration, 2024), obscuring the correct perception of events in the case of two or three simultaneous transits.

The machine learning side

In this section of the project, the application of Machine Learning (ML) concepts in the search for exoplanets through transit photometry will be discussed. Mitchell (1997) provides a succinct definition of ML: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E". Therefore, the fundamentals of ML theory can be categorised into three sections: types of experiences, types of tasks, and types of performance measures. AI and ML share similarities, as ML constitutes a significant branch of AI (IBM, 2023). In the context of exoplanet detection, the terms AI and ML are used interchangeably because AI systems in this field heavily rely on ML.

Types of ML tasks and performance measures and how they differ

There are several tasks that ML can undertake. The primary ones encompass classification (e.g., sorting emails into 'spam' or 'not spam'), regression (predicting house prices), anomaly detection (spotting fraudulent credit card activity), and clustering (grouping customers by their buying patterns) (Goodfellow et al., 2016, p.100; Kapitanova et al., 2012). For transit photometry, classification emerges as the most critical task, as it involves determining whether a star hosts an exoplanet. Thus, each star is typically classified into two categories: either having one or more exoplanets or having none. Consequently, software executing classification is tasked with assigning the given input to one of n categories (Goodfellow et al., 2016, p.98). In regression, a program is tasked with predicting a numerical value based on specific input parameters (Goodfellow et al., 2016, p.99). For example, regression can forecast the size of an exoplanet based on the characteristics of a star and the luminosity-time graph. Anomaly detection involves sifting through a series of events or objects and identifying some as uncommon or abnormal

(Goodfellow et al., 2016, p.100). When a computer program performs clustering, it partitions a dataset into subgroups or clusters based on similarity (Jung, 2022). For instance, ML algorithms might group exoplanets based on their orbital periods. Thus, ML exhibits versatility in performing various tasks. Each task may prove beneficial in certain scenarios but less so in others, hence the selection of specific tasks depends on the field of study and research objectives.

When assessing the performance of a model, there typically exists a specific quantitative measurement method for each task assigned to the model (Goodfellow et al., 2016, p.102). For instance, the performance of classification models is evaluated using metrics such as accuracy or error rate (Goodfellow et al., 2016, p.102). These metrics are inversely related, as accuracy quantifies the proportion of correct outputs, while error rate quantifies the proportion of incorrect outputs. Primarily, a typical concern lies in evaluating how well the ML algorithm performs on unseen data, as this reflects its effectiveness in real-world deployment. Consequently, these performance indicators are assessed using a separate test dataset, independent of the data used for training the ML system (the dataset utilised for learning is termed a training dataset, while the dataset employed to evaluate how effectively the model learned from the training dataset is referred to as the test dataset). It is important to note that whilst the idea of selecting performance measures may seem obvious and clear, it often proves challenging to identify a performance measure that correlates closely with the anticipated behaviour of the model or program (Goodfellow et al., 2016, p.102).

Types of ML experiences and how they differ

ML techniques analyse data points generated within a specific application domain. Each data point possesses distinct attributes, which can be categorised into two groups: features and labels (Goodfellow et al., 2016, p.103). Features are quantifiable attributes that can be automatically measured and computed, while labels represent higher-level truths or quantities of interest that are not easily measurable and often require human expertise for discovery. ML algorithms are typically classified into three categories: supervised, unsupervised, and reinforcement, based on the type of experience they are exposed to throughout the learning phase (Goodfellow et al., 2016, p.104).

Firstly, supervised learning methods depend on a training set comprising labelled data points with known label values (Jung, 2022). Human experts assign values to data points based on their characteristics (Jung, 2022). For instance, if a model was to be trained to identify whether there is a cat on a picture, then the labelling would be done by classifying each existing image as 'having a cat' or 'not having a cat.' Supervised ML entails fitting a curve with an equation to the labelled data points in training sets (Kapitanova et al., 2012). To perform curve fitting, a loss function is required to quantify the fitting error. Various supervised ML techniques employ multiple loss functions to quantify the difference between data points' actual and predicted labels (Jung, 2022). The curve's function is termed a classifier for discrete outputs, whereas a regression function for continuous outputs (Kapitanova et al., 2012).

Secondly, we turn to unsupervised learning. As the term suggests, this method lacks supervision and therefore does not rely on labelled data. In many instances, the volume of data is extensive, making it impractical to label every training data point accurately due to the time and effort involved. For example, consider the scenario where a ML algorithm is tasked with identifying animals in pictures. Numerous images sourced from the internet serve as input, but most of them are not labelled; they are simply posted by various users. In unsupervised learning, the algorithm receives input data that lacks labels. Its objective is to discern patterns within the data to make accurate predictions for new instances (Kapitanova et al., 2012).

The third type is reinforcement learning, whose meaning may initially appear unclear. In this form of ML experience, algorithms do not rely on an established dataset but instead engage with their environment, creating a feedback loop between the learning system and its experiences

(Goodfellow et al., 2016, p.104). For example, it could be a mail delivery robot put into a city and it now learns how to move and what to do. Through a series of rewards and punishments, it figures out how to deliver mails. However, in most instances, a dataset is available, rendering reinforcement learning less common (Goodfellow et al., 2016, p.104).

In the context of transit photometry, supervised learning is predominantly employed. This is because algorithms are trained using confirmed exoplanets that have been scrutinised using various techniques (not solely transit photometry), with unlabelled data subsequently used for testing potential discoveries.

Neural networks and why they are beneficial

Given that many cutting-edge technologies in astronomy rely on Neural Networks (NNs) and Deep Neural Networks (DNNs), an introduction to this topic is warranted. NNs have found application in various planetary science tasks, such as multi-planet prediction and atmosphere categorisation (Pearson et al., 2018). Furthermore, the subsequent section will demonstrate that the majority of researchers employing ML for these purposes utilise DNNs.

According to Aggarwal (2018), understanding neural networks begins with an analogy to the human learning process. In humans, neurons, the basic units of the nervous system, communicate via axons and dendrites, with synapses forming connections between them. Artificial neural networks mirror this biological mechanism, comprising computing units called neurons. Each neuron receives input adjusted by a weight, influencing its function. Inputs are processed in neural networks, with calculated values transmitted through neurons, utilising weights as interim parameters. Learning occurs through weight adjustments based on training data, which includes input-output pairs representing the function to be learned. For example, pixel images and corresponding labelled annotations (e.g., fox) serve as training data. The network uses input representations to predict output labels, and training data assesses the accuracy of predictions.

Errors prompt adjustments in synaptic strengths, akin to learning in biological organisms. Similarly, neural network weights are modified in response to prediction errors, refining the computed function and enhancing prediction accuracy. By adjusting weights across numerous input-output pairs, the network evolves, leading to more precise predictions. Thus, exposure to diverse images of foxes during training enables the network to identify unseen images of them accurately.

Putting it simply, imagine neural networks like interconnected switches or gates. Each gate receives some information, does a quick calculation to decide how important it is (weight), and passes it along to the next gate, helping the network learn patterns.

When considering deep learning neural networks, DNNs consist of a NN composed of dozens to millions of interconnected neurons (Jung, 2022). Notably, deeper neural networks necessitate fewer units per layer owing to the potent composition functions offered by successive levels (Aggarwal, 2018). Figure 2 illustrates the structural disparity between a NN and a DNN (Kumar, 2018).

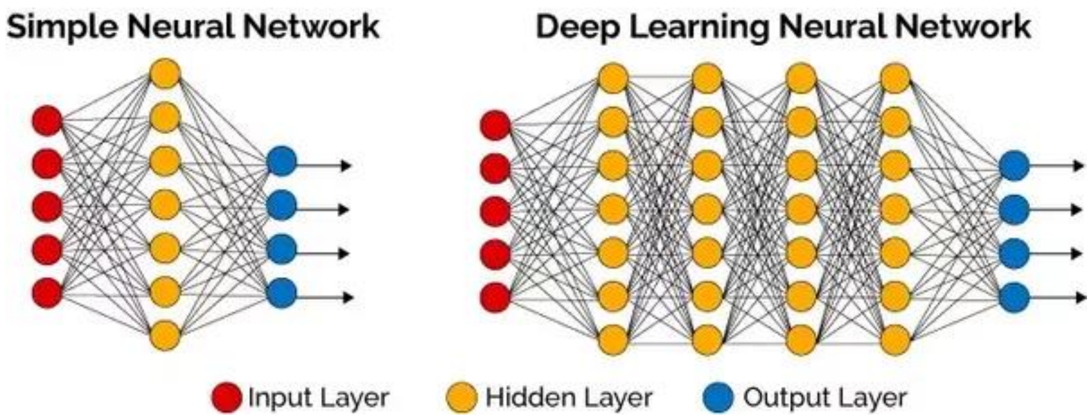

Figure 2: Difference between a NN and DNN (Source: Kumar, 2018)

Additionally, there exists a type of DNN known as a Convolutional Neural Network (CNN), tailored for processing input with a predefined grid-like structure (Goodfellow et al., 2016, p.326). Examples include time-series data, represented as a 1-D grid collecting samples at regular intervals, and image data, represented as a 2-D grid of pixels (Goodfellow et al., 2016, p.326). Convolutional networks have demonstrated effectiveness in practical applications, particularly in computer vision for tasks such as object recognition and classification (Aggarwal, 2018). Structurally, CNNs feature additional convolution layers, where a convolution operation is defined (Aggarwal, 2018). Convolution can be understood as a specific type of linear operation (Goodfellow et al., 2016, p.326), potentially useful for exoplanet identification via photometry due to regular brightness measurements. DNNs excel at recognizing small features within vast datasets, rendering them advantageous (Pearson et al., 2018). These algorithms optimise weights to minimise the disparity between the deep net output and the training data's expected value, facilitating effective learning (Pearson et al., 2018).

Common problems and disadvantages related to ML application

No ML model is flawless; they inevitably exhibit drawbacks and encounter common issues. Firstly, overfitting is a prevalent concern. Merely training a model on a dataset does not ensure optimal performance on unseen test data, even if it accurately predicts target values within the training dataset (Aggarwal, 2018). Consequently, a disparity between training and test performance always exists, particularly pronounced when models are complex and datasets are sparse. Regularisation techniques are necessary to mitigate overfitting.

Secondly, training data may require extensive periods. It is commonplace for a NN to necessitate several hours or even days to complete training. However, strategies exist to address this issue, such as partitioning the input data into distinct subsets and allocating each chunk to a separate processing unit (Agarwal, 2022).

Finally, when employing a NN, monitoring its progress becomes nearly impossible. It is not feasible to peer directly into a deep neural network to comprehend its functioning (Knight, 2017). A network's rationale lies within the operation of thousands of simulated neurons, arranged into numerous interconnected layers (Knight, 2017). Regarding the accuracy of ML techniques, Figure 3 illustrates a comparison of the accuracies of standard ML models and Deep NNs (Aggarwal, 2018).

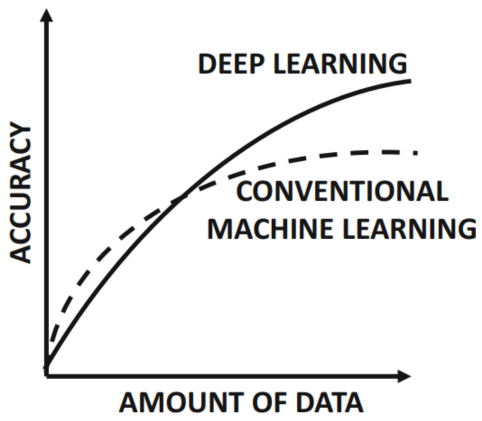

Figure 3: Accuracy comparison between traditional ML and DNN (Source: Aggarwal, 2018)

As depicted in the graph, conventional ML models outperform DNNs when the dataset is limited in size. However, DNNs significantly outshine typical ML methods when presented with extensive datasets due to the availability of sufficient data and/or computational power (Aggarwal, 2018). Therefore, the selection of an appropriate ML method should be contingent upon the available data.

Analysis of various studies on using ML for this purpose

This section will analyse and synthesise several studies on the utilisation of ML in the search for exoplanets through the transit photometry method. By meticulously examining each research paper (and citing the authors at the beginning of each paragraph to indicate that all information in that paragraph is adapted from that source, thereby avoiding over-citation), I aim to recognise the ways ML contributes to the detection of new exoplanets. To provide a comprehensive comparison between contemporary models and traditional ones, it is crucial to first introduce a pair of models that do not rely on ML.

Hippke and Heller (2019) employed the transit least squares (TLS) algorithm, grounded in mathematical principles, to delineate the transit graph's behaviour. The researchers actively juxtaposed their algorithm with a similar, albeit older, model known as box least squares (BLS), conventionally utilised for planetary transit searches within extensive datasets. In the BLS technique, the transit photometric profile is modelled in the form of a negative boxcar function that has a default average zero flux outside of transit and a constant drop throughout the transit. Such a methodology is the reason why this model is so computationally efficient and it also facilitates precise identification of signals with signal-to-noise ratios (S/N) ranging from medium to high, encompassing Jupiter-sized and Neptune-sized planets revolving around stars that are similar to the Sun in surveys. However, BLS has a systematic noise factor due to the nature of the boxcar function. The two-state model with constant out-of-transit and in-transit flux disregards stellar limb darkening and planetary ingress/egress in the light curve, thus generating supplementary noise in the test statistic, leading to a decrease in low-S/N signals. In contrast, TLS aims to mitigate this systematic noise component in the search statistics. Initially, TLS selects a function for a curve that best represents existing planetary transit light curves, serving as the standard template for future exoplanet identifications to maximise efficiency. This program adopts a predefined parameterisation to assume a practical transit shape, encompassing ingress, egress, and star limb dimming, derived from prior planetary detections. *In simple terms*, the TLS and BLS methods both try to find exoplanets by looking at how stars' brightness changes. TLS is newer and considers more realistic star brightness variations, while BLS uses a simpler model. Figure 4 depicts a curve fitting conducted by the two models, showing their distinction (Wells, 2020).

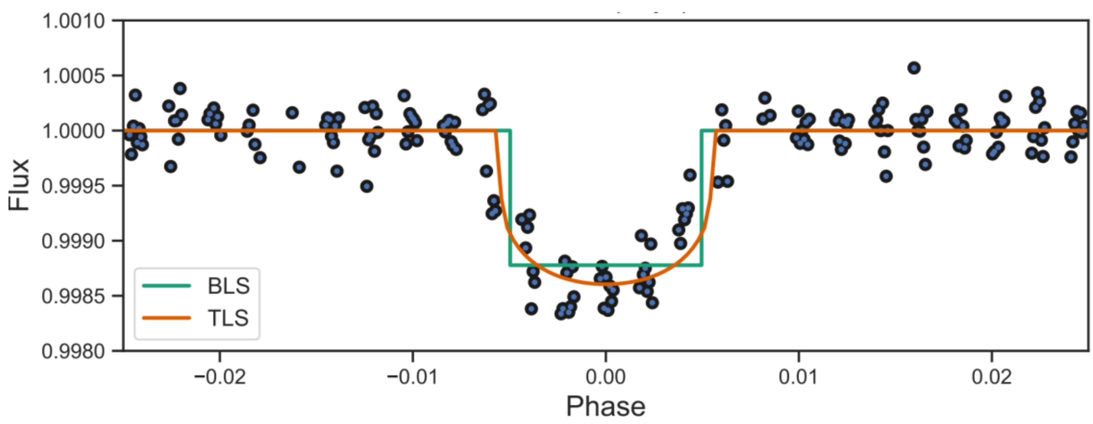

Figure 4: TLS and BLS in use for a light curve of K2-3 b. TLS has a more representative line of best fit of the transit (Source: Wells, 2020)

Hippke and Heller (2019) refined the model primarily for detecting small planets, yet the algorithm has exhibited superior performance over BLS in detecting large planets. The TLS technique aims to identify periodic transit-shaped signals within time series flux measurements. The technique functions by folding the data across a variety of trial periods (selected duration of one cycle around the orbit), transit epochs (start time/external ingress), and durations of transits, subsequently computing the statistic of the phase-folded light curve considering a specific number of data points in the appropriate model and the recorder numbers. Thus, for every trial period, TLS seeks the minimum statistic variable by exhaustively trying all pairings of transit durations and epochs. Consequently, TLS surpasses BLS in detecting various types of planets, particularly grazing transiters and very small planets. Furthermore, for small planets, TLS demonstrated a 17% increase in true positives (~76% for BLS and ~93% for TLS) with appropriately configured parameters. Regarding the comparison with ML models, the authors note that while ML algorithms such as deep learning and Random Forest outperform the BLS model, the computational costs for training the model are notably high. Additionally, understanding the origin of results in the case of DNNs is challenging due to numerous hidden layers. It is pertinent to mention that the paper was authored in 2019, which may be outdated concerning the discussion of ML success. The article briefly considers some studies suggesting that random forest classifiers and CNNs exhibit a high number of false-positives and more false-negatives than BLS.

Nevertheless, the researchers acknowledge the necessity for an independent benchmark to enable a fair comparison between ML models and BLS and TLS.

Pearson et al. (2018) centred their investigation on the application of AI in the quest for exoplanets by creating a neural network. The training dataset for the DNNs comprised 311,000 samples of both transit and non-transit events. Then, the researchers tested their DNNs using light curves produced during the Kepler mission, where the model had to identify periodic transits of exoplanets that matched the correct period, not requiring any curve fitting. In addition, different strategies were implemented to enhance planet detection rates such as 1D convolutional networks and feature modifications like wavelets. A CNN demonstrated significant improvement, outperforming the conventional BLS method. During testing, BLS achieved a true positive rate of 75%, whereas a 1D CNN model demonstrated 96%. The researchers concluded that machine learning can effectively detect subtle characteristics from large datasets, surpassing human performance, and holds potential for the future of analysis in astronomy.

Similarly, Zucker and Giryes (2018) utilised CNNs in their study to validate the potential of ML tools. They acknowledged the challenge of identifying exoplanets in solar-like star systems, attributed to the existence of red noise in light curves collected from specialised space telescopes and the shallow and brief nature of transits. The researchers trained their CNN on simulated data and subsequently applied it to a simulated dataset. Their findings suggested that even highly complex cases could potentially be detected with a satisfactory degree of accuracy. Regarding model comparison, the CNN was pitted against the traditional BLS model, albeit with slight modifications to ensure a fair comparison, with the CNN demonstrating superior performance once again, underscoring the dominance of machine learning. Figure 5 illustrates one measurement of comparing it to the BLS model.

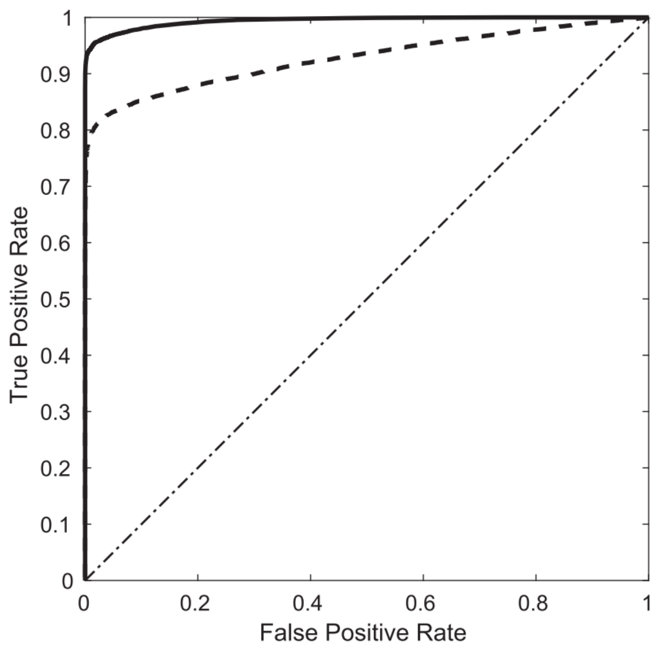

Figure 5: Receiver Operating Characteristic (ROC) curve comparing the neural network model's performance (solid line) with the Box Least Squares (BLS) method enhanced by a high-pass filter (dashed line). The dotted-dashed line indicates random chance performance ("no-discrimination" line).

The Figure 5 clearly suggests that the neural network significantly outperforms the BLS method, as its curve is closer to the top-left corner, indicating a higher true positive rate and lower false positive rate. The scientists observed that the model adeptly handled outlier removal and discontinuity detection, tasks that might have necessitated an additional stage otherwise. Finally, the researchers compared their study with the one done by Pearson et al. (2018). The two papers adopted fundamentally different approaches to simulate noise. Zucker and Giryes (2018) argued that Pearson et al.'s method made the data highly unrealistic. Nevertheless, Zucker and Giryes (2018) acknowledged the importance of Pearson et al.'s paper as a significant milestone contributing to the field.

A more recent paper authored by Pearson (2019) directed its attention towards discerning complex multiplanet systems using data sourced from the TESS mission. The researcher employed advanced mathematical techniques and simulations such as linear orbit ephemeris (a formula used to predict exactly when a planet will transit its star), Bayesian evidence (which helps determine how likely one model is compared to another based on observed data), N-body simulation (computer models that predict how multiple objects in space move and interact), nested sampling (a statistical method used to compare how well different models explain observed data), and ultimately a ML model to ascertain whether the orbits of potential planets are influenced by other planets, utilising AI to facilitate predictions concerning these potential planets. A dual-input, multi-output CNN with an accuracy of 99% at S/N > 2 constituted the final stage of the investigation. As for the findings, the model successfully identified three multiplanet systems, some of whose planets did not directly transit in front of Earth, warranting further scrutiny. This demonstrates the feasibility of meticulously studying and resolving such intricate scenarios. Pearson acknowledges the model's current inability to distinguish between eclipsing binaries (binaries are systems involving two stars interacting with each other) and transit signals, a deficiency to be addressed in future iterations of the AI model.

In a similar vein of identifying multiplanet systems, a study conducted by Yeh and Jiang (2021) delves into the application of CNNs for analysing light curves obtained from the BRITE satellite. Scientists underscored the significant advantage that ML holds over traditional methods, owing to the inherent self-improvement and self-modification characteristics of NNs, with CNNs being actively employed in the pursuit of exoplanet discovery. Researchers utilised ML to detect exoplanet transits using photometric data from the BRITE spacecraft. Four CNN models were deployed for exoplanets with orbital periods spanning from one to five days, with each interval being addressed by a separate model. This partitioning of intervals may aid in identifying complex cases of multiple planets within extrasolar systems. The models were trained using synthetic data combined with the BRITE light curves. Upon testing the models on 35 light curves of stars observed by BRITE, the results revealed that 10 of them were deemed potentially harbouring an exoplanet by at least one CNN model, with only two of those 10 being identified by two or more models, suggesting a high likelihood of exoplanet presence. High-precision observational data will be instrumental in confirming the existence of exoplanets in these 10 candidates and refining orbital parameter determination. While the CNN algorithms demonstrated success in identifying potential transit candidates, further observational data is imperative for their validation. Despite employing different methodologies, both Pearson (2019) and Yeh and Jiang (2021) achieved noteworthy success in their respective endeavours.

Innovative research by Wang et al. (2024) presents the GPFC, or Graphics Processing Unit (GPU) Phase Folding and CNN, approach for identifying transit signals and assesses its effectiveness in detecting periodic transits in star photometric time series. The phrase “phase folding” means aligning repeating signals from stars to more easily detect planets. The model integrates a GPU phase-folding program that has a CNN to detect signals of transit from exoplanets of small size using unprocessed photometry profiles. This technique is successful because GPUs efficiently process data because they can handle many calculations simultaneously, greatly speeding up tasks like analysing large datasets in astronomy. The researchers utilised over two million imitated light curves with distributions of the Kepler parameters to compare the precision and effectiveness of the GPFC model with the BLS approach. Their analysis specifically focuses on cases with low S/N ratios, wherein GPFC outperforms BLS. Figure 6 illustrates the results of one of the metrics tested.

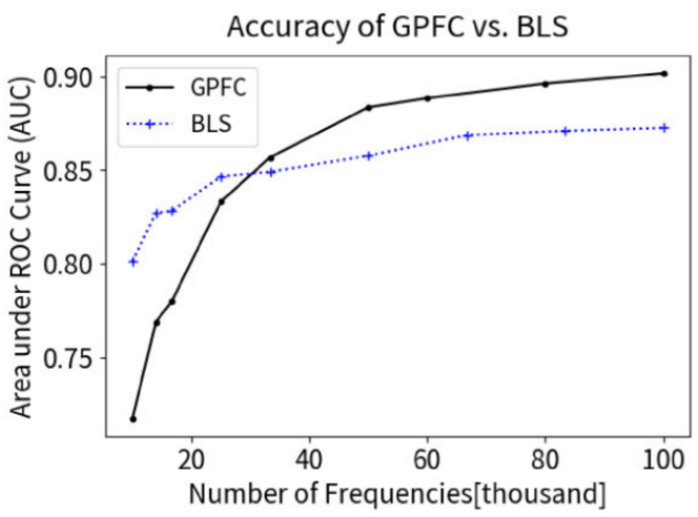

Figure 6: A comparison of the accuracies of GPFC and BLS as the number of trial frequencies alters. In this figure, accuracy indicates the ability to differentiate between transit and non-transit light curves, and it is represented by the area under the receiver operating characteristic (ROC) curve. (Source: Wang et al., 2024)

Figure 6 shows ROC curves, which compare how well AI models and traditional methods detect exoplanets. Higher curves mean better performance, clearly showing that AI methods consistently outperform traditional methods. The modern technique correctly identified 100% of the confirmed exoplanets from the Kepler mission while it was searching for exoplanets with a period range from 0.2 to 1.0 days. Consequently, the new model demonstrated a considerable advantage in accuracy, alongside approximately 1000 times greater computational speed. Notably, they optimised the traditional BLS algorithm, resulting in significant speed enhancements, although it remained approximately 15 times slower than GPFC. Lastly, GPFC has shown to have the ability to swiftly discover novel, subtle transit signals in light curves from the Kepler mission, demonstrating its potential utility in extensive transit surveys in the future.

In a study conducted by Tey et al. (2023), researchers utilised the pre-existing neural network ‘Astronet-Triage-v2’ to analyse data from the TESS mission, categorising light curves into five groups: ‘periodic eclipsing signal’, ‘single transit’, ‘contact eclipsing binaries’, ‘junk’, and ‘not sure’. The evaluation of the model involved two key metrics: recall (the ratio of number of true positive datapoints to the total number of initial datapoints with positive labels) and precision (the ratio of number of true positive datapoints to total number of datapoints predicted as positive), commonly employed in ML contexts. Through testing, the researchers determined a recall rate of approximately 97% and a precision rate of around 76%. Furthermore, they compared the performance of the model with an earlier version, ‘Astronet-Triage,’ which identified 200 fewer exoplanets out of just over 4000, indicating a significant divergence in success between the two AI models tested. Interestingly, the researchers delved into the factors limiting the accuracy of their model. They identified binaries as a source of confusion, blurring the distinction between two categories that ML aims to differentiate. Additionally, noisy data poses a significant challenge, complicating the detection process by making it harder to distinguish between junk and genuine transit signals. In contrast, mathematical models would be inadequate for such complex classification tasks, as they merely fit a template curve to measure its resemblance to the light curve graph.

Finally, Mislis, Pyrzas and Alsubai (2018) devised an unsupervised ML model named TSARDI with the objective of enhancing transit detection efficacy in large-scale survey data. In contrast to prior models reliant on NNs, TSARDI employs the DBSCAN clustering algorithm, serving as a robust filter to identify and eliminate residual noise points following detrending procedures. This approach operates autonomously (as an unsupervised model) on individual light curves, necessitating no prior knowledge of other field light curves. By performing simulated transit searches with artificial planetary signals added to real data from the Qatar Exoplanet Survey project, TSARDI demonstrated an approximately 11% improvement in overall transit identification efficiency compared to the conventional sigma-clip algorithm used for outlier removal by repeatedly excluding values that deviate significantly from the average. Additionally, for brighter host stars (magnitude < 12), TSARDI achieved a detection efficiency of approximately 80% for simulated planets. Although initially developed for exoplanet detection, TSARDI holds the potential for adaptation in various time series analyses. Whilst this study does not include a comparison with other models beyond a conventional sigma-clip algorithm, it introduces an intriguing new approach utilising unsupervised ML to enhance efficiency.

The role of AI in detecting exoplanets via the transit photometry method

While it is feasible to achieve satisfactory results without ML, for instance, by employing the BLS model, the utilisation of NNs provides superior accuracy within a shorter time frame, making them an indisputably dominant choice. Despite the existence of other viable ML algorithms, such as clustering, their accuracy has been demonstrated to be inferior. ML typically manages the task autonomously, with human intervention limited primarily to data labelling for supervised ML, whereas mathematical models such as BLS are initially crafted manually.

Furthermore, the comparisons and analyses conducted between ML and non-ML models, as well as among different ML approaches, indicate that ML offers numerous advantages in the realm of exoplanet detection using transit photometry. Firstly, ML substantially reduces processing time owing to its computational prowess. It can swiftly handle vast datasets in comparison to human labour and even non-ML mathematical models. Secondly, ML-powered models are exceedingly commendable in recognising subtle features in extensive datasets, proving particularly beneficial in intricate cases depicted in light curve graphs. Thirdly, neural networks exhibit self-modification capabilities, obviating the need for human intervention. ML has found widespread application in planet identification and astrophysics in general, with CNNs being highlighted as pivotal components of successful research, according to Yeh and Jiang (2021) and the results of the papers considered. Fourthly, ML-powered models surpass other types of models in exoplanet identification via transit photometry. It has been observed that scientists consistently achieve much higher accuracy outputs using CNNs compared to BLS and TLS. Table 1 summarises my findings of the academic articles.

Research paper: the model used	**Main Comparison Result / Outcome**
Hippke and Heller (2019): TLS model	True positive rate of BLS and TLS: 76% vs 93%
Pearson et al. (2018): 1D CNN model	True positive rate of BLS vs 1D CNN: 75% vs 96%
Zucker and Giryes (2018): CNN model	The ROC curve is significantly better compared to BLS (Figure 5).
Pearson (2019): dual-input, multi-output CNN	Accuracy of 99% at S/N > 2.
Yeh and Jiang (2021): multiple CNNs	Out of 35 stars, two star systems are confidently identified as having exoplanets
Wang et al. (2024): the GPFC model	Correctly identified 100% of the confirmed exoplanets from the Kepler mission. Considerable advantage in accuracy, while being many times faster than BLS (Figure 6).
Tey et al. (2023): Astronet-Triage-v2	Recall rate of around 97%. Precision rate of around 76%.
Mislis, Pyrzas and Alsubai (2018): TSARDI	Detection efficiency of about 80% for simulated planets.

Table 1: Summary of synthesising various academic papers on ML in exoplanet detection via transit photometry.

Although there is no direct comparison found between ML-based models and TLS, and while numerous comparisons against BLS have been made, the available information suggests that the true positive rate for TLS at specific settings is 93%, which is lower than the general value for ML models. This lack of comparison to a more advanced mathematical model, TLS, can be considered a potential limitation of the project. Despite this, ML models are dominant over BLS and, by an accepted assumption which is based on one data point of the accuracy of TLS, ML is more successful than TLS. Moreover, the first two rows in Table 1 (Hippke and Heller, 2019, and Pearson et al, 2018) provide almost identical tests of performance, since 75% and 76% are extremely close. Therefore, it is valid to approximate the results and state that the True Positive rate of TLS vs 1D CNN is around 93% vs 96%. This justifies my assumption and confirms the ML's dominance. These two factors and the results from Table 1 indicate that ML is superior to conventional methods. Lastly, ML models offer greater flexibility as a rapidly evolving field that can be enhanced in various aspects such as computation speed, adaptability, and methodology. ML models are highly generalisable, implying that a model developed for exoplanet detection could be easily modified and applied to another field, as was mentioned in a few research papers.

However, ML does entail several drawbacks, which are undoubtedly outweighed by the preceding benefits. These include the risk of overfitting, the substantial time required for model training, and the lack of control over the training process of a NN.

This reasoning leads to the conclusion that there is ample evidence supporting the efficacy and tremendous utility of ML, as seen in the results in Table 1. It significantly expedites data processing, enhances accuracy, and offers additional advantages. Therefore, it would be reasonable and prudent to assert that AI currently plays a crucial role in identifying exoplanets using the transit photometry method and aids scientists in their endeavours to a great extent. Additionally, ML can synergistically collaborate with other potent methods to achieve desired outcomes while serving as a fundamental component.

Conclusion

One of the most profound mysteries facing humanity is the existence of other life forms in space, inhabiting planets within distant star systems. Recently, AI has emerged as a valuable tool for automating processes and managing vast datasets, with one of its current applications being in the field of astronomy. In this study, the focus was on the use of ML for detecting exoplanets through the transit photometry method, aiming to assess its helpfulness in this context. The investigation commenced with an examination of the underlying physics principles behind the transit photometry technique, followed by a discussion of its advantages and limitations. Subsequently, attention turned to the ML aspect of the process, which involves analysing light curves and classifying whether a particular curve represents an exoplanet using a specific AI model. DNNs have gained widespread popularity due to their high level of accuracy and capability to handle complexity. A comprehensive analysis of various studies pertaining to this topic revealed that ML models, especially CNNs, outperform traditional methods. The comparison between the conventional method, BLS, and modern ML models consistently showed ML to be significantly more successful. ML has emerged as the preferred option due to its numerous benefits, including high computational speed, the ability to identify subtle features, self-improvement of neural networks, enhanced accuracy, and adaptability to diverse tasks. Consequently, AI now plays a pivotal role in exoplanet detection through the transit photometry technique, substantially assisting scientists in their pursuits and capable of complementing other methodologies. The study successfully addressed all aims and objectives, with certain aspects, such as neural networks, being explored in considerable depth. The research question was confidently answered, considering findings from the scientific papers investigated. However, further examination of the newly proposed mathematical model, TLS, is warranted to assess its potential accurately. In this study, TLS capabilities were approximated based on limited data, leading to the conclusion that, as BLS, it is surpassed by neural networks. Moreover, this research focused solely on the most prominent and successful method of exoplanet detection, suggesting that future research exploring alternative techniques may yield different conclusions. The deployment of ML for exoplanet identification has unequivocally demonstrated its significant advantages. Given ML's continual potential for advancement, it is poised to emerge as an increasingly reliable and indispensable tool for scientists, holding immense promise for driving future breakthroughs in astronomy.

Bibliography

Agarwal, B. (2022). *How to Reduce the Training Time of Your Neural Network from Hours to Minutes.* Available at: https://towardsdatascience.com/how-to-reduce-the-training-time-of-your-neural-network-from-hours-to-minutes-fe7533a3eec5 (Accessed: 8 February 2024).

Aggarwal, C. C. (2018). *Neural Networks and Deep Learning.* s.l.:Springer International Publishing.

American Museum of Natural History (2024). *Guided Exploration: Stars Zone.* Available at: https://www.amnh.org/exhibitions/permanent/the-universe/educator-resources/guided-exploration-stars-zone (Accessed: 19 January 2024).

Balter, A. (2022). *What is Astrophysics?.* Available at: https://www.space.com/26218-astrophysics.html (Accessed: 7 December 2023).

Barberio, E. (2020). *Exploring The Most Unknown Universe.* Available at: https://pursuit.unimelb.edu.au/articles/exploring-the-most-unknown-universe (Accessed: 7 December 2023).

British Astronomical Association (2022). *Exoplanet Transit Imaging and analysis Process.* Available at: https://britastro.org/section_information_/exoplanets-section-overview/exoplanet-transit-imaging-and-analysis-process (Accessed: 19 January 2024).

European Space Agency (2019). *How to find an exoplanet.* Available at: https://www.esa.int/Science_Exploration/Space_Science/Exoplanets/How_to_find_an_exoplanet (Accessed: 22 December 2023).

Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep Learning*. s.l.:MIT Press.

Gotame, R. C. (2020). *Transit Photometry Method for Finding the Exoplanets.* Available at: https://physicsfeed.com/post/transit-photometry-method-finding-exoplanets/ (Accessed: 23 December 2023).

Hippke, M. and Heller, R. (2019). 'Optimized transit detection algorithm to search for periodic transits of small planets', *Astronomy&Astrophysics*, 623, pp.1-13. doi: https://doi.org/10.1051/0004-6361/201834672

IBM (2023). *What is machine learning?.* Available at: https://www.ibm.com/topics/machine-learning (Accessed: 5 January 2024).

Jung, A. (2022). *Machine Learning: The Basics.* Singapore: Springer.

Kapitanova, K., Son, S., Hu, F. (ed) and Hao, Q. (ed) (2012). *Intelligent Sensor Networks.* New York: CRC Press.

Knight, W. (2017). *The Dark Secret at the Heart of AI.* Available at: https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/ (Accessed: 6 February 2024).

Kumar, H. (2018). *How deep should neural nets be?.* Available at: https://kharshit.github.io/blog/2018/04/27/how-deep-should-neual-nets-be (Accessed: 28 January 2024).

Lissauer, J. J. (2024). *Extrasolar planet.* Available at: https://www.britannica.com/science/extrasolar-planet (Accessed: 17 December 2023).

Mislis, D., Pyrzas, S. and Alsubai, A. (2018). 'TSARDI: a Machine Learning data rejection algorithm for transiting exoplanet light curves'. *Monthly Notices of the Royal Astronomical Society,* 481(2), pp.1624-1630. doi: https://doi.org/10.1093/mnras/sty2361

MIT and the Centre for Astrophysics (2018). *TESS Mission.* Available at: https://tess.mit.edu/science/ (Accessed: 18 December 2023).

Mitchell, T. M. (1997). *Machine Learning.* New York: McGraw-Hill.

NASA Exoplanet Exploration (2024). *5 Ways to Find a Planet.* Available at: https://exoplanets.nasa.gov/alien-worlds/ways-to-find-a-planet/ (Accessed: 18 December 2023).

Pearson, K. A. (2019). 'A Search for Multiplanet Systems with TESS Using a Bayesian N-body Retrieval and Machine Learning'. *The Astronomical Journal,* 158(6), pp.243-260. doi: https://doi.org/10.3847/1538-3881/ab4e1c

Pearson, K. A., Palafox, L. and Griffith, C. A. (2018). 'Searching for exoplanets using artificial intelligence'. *Monthly Notices of the Royal Astronomical Society,* 474(1), pp.478-491. doi: https://doi.org/10.1093/mnras/stx2761

Santerne, A. et al. (2012). 'SOPHIE velocimetry of Kepler transit candidates'. *Astronomy&Astrophysics,* 545, pp.1-16. doi: https://doi.org/10.1051/0004-6361/201219608

Seager, S. (2015). 'Searching for Other Earths'. *The New Atlantis,* 47, pp.67-75. https://www.jstor.org/stable/43671541

Tey, E. et al. (2023). 'Identifying Exoplanets with Deep Learning. V. Improved Light-curve Classification for TESS Full-frame Image Observations'. *The Astronomical Journal*, 165(3), pp. 165-195. doi: https://doi.org/10.3847/1538-3881/acad85

The International Astronomical Union (2018). *Pluto and the Developing Landscape of Our Solar System.* Available at: https://www.iau.org/public/themes/pluto/ (Accessed: 17 December 2023).

The Planetary Society (2020). *Down in Front!: The Transit Photometry Method.* Available at: https://www.planetary.org/articles/down-in-front-the-transit-photometry-method (Accessed: 22 December 2023).

The Planetary Society, n.d. *Exoplanets, worlds orbiting other stars.* Available at: https://www.planetary.org/worlds/exoplanets (Accessed: 17 December 2023).

Wang, K., Ge, J., Willis, K. and Zhao, Y. (2024). 'The GPU phase folding and deep learning method for detecting exoplanet transits. *Monthly Notices of the Royal Astronomical Society*, 528(3), pp.4053-4067. doi: https://doi.org/10.1093/mnras/stae245

Webb Space Telescope (2024). *About the Telescope.* Available at: https://webbtelescope.org/about (Accessed: 18 December 2023).

Wells, R. (2020). *The properties of transiting exoplanets*. Belfast: Queen's University Belfast.

Wenz, J. (2023). *How the first exoplanets were discovered.* Available at: https://www.astronomy.com/science/how-the-first-exoplanets-were-discovered/ (Accessed: 18 December 2023).

Williams, M. (2017). *What is the Transit Method?.* Available at: https://www.universetoday.com/137480/what-is-the-transit-method/ (Accessed: 23 December 2023).

Winn, J. (2023). *The Little Book of Exoplanets.* s.l.:Princeton University Press.

Yeh, L.-C. and Jiang, I.-G. (2021). 'Searching for Possible Exoplanet Transits from BRITE Data through a Machine Learning Technique'. *The Astronomical Society of the Pacific*, 133, pp.1-12. doi: https://doi.org/10.1088/1538-3873/abbb24

Zucker, S. and Giryes, R. (2018). 'Shallow Transits—Deep Learning. I. Feasibility Study of Deep Learning to Detect'. *The American Astronomical Society*, 155, pp.147-155. doi: https://doi.org/10.3847/1538-3881/aaae05

Literature review

Source	Evaluation
Agarwal, B. (2022). *How to Reduce the Training Time of Your Neural Network from Hours to Minutes.* Available at: https://towardsdatascience.com/how-to-reduce-the-training-time-of-your-neural-network-from-hours-to-minutes-fe7533a3eec5 (Accessed: 8 February 2024).	**Currency:** Good – quite recent and completely acceptable. Additionally, the article is about NNs, and by 2022, the topic was known and studied enough to state that the article is absolutely applicable today. **Relevance:** Very relevant to my point of stating that NNs require a lot of time to train because the article discusses it and the solution to the issue. **Authority:** The article was accepted and published on a well-known and reputable website 'Towards Data Science', which is primarily about ML. The person who wrote the article is a data scientist and astrophysicist, which means that he is an expert in this field and the information is reliable. **Accuracy:** No crucial inaccuracies. The article is high-quality and explains the issue in detail. Everything is based on facts and is strictly scientifically proven.
Aggarwal, C. C. (2018). *Neural Networks and Deep Learning.* s.l.:Springer International Publishing.	**Currency:** 2018 - not too old, but there may be some new technologies as the field of ML advances. Nevertheless, the book covers the basics, which remain the same. **Relevance:** It is directly related to my project as I was describing the origin of NNs and some other fundamental ideas. **Authority:** Charu C. Aggarwal has an excellent education, and currently works as a distinguished research staff member at IBM, a leading computer science company. He has written 10 books, and published over 400 papers in recognized journals. His h-index is 138, which is truly extraordinary. **Accuracy:** The book is high-quality and very reliable. Everything is based on facts and is strictly scientifically proven.
Barberio, E. (2020). *Exploring The Most Unknown Universe.* Available at: https://pursuit.unimelb.edu.au/articles/exploring-the-most-unknown-universe (Accessed: 7 December 2023).	**Currency:** Not the most recent article for the fast-paced field of astrophysics, but it is also not very old, therefore it is somewhat acceptable. **Relevance:** It is loosely relevant to my essay because I was introducing the vastness and mystery of our universe by talking about dark matter and energy. **Authority:** The article was published by a professor Elisabetta Barberio from the University of Melbourne, which is considered a world-class university. The professor has 2039 scholarly works, which indicates the high expertise she possesses. **Accuracy:** No obvious inaccuracies. The article is high-quality and explains the issue in detail. Everything is based on facts and was verified and accepted by the University of Melbourne, indicating a high level of reliability.
British Astronomical Association (2022). *Exoplanet Transit*	**Currency:** 2022 – very recent, taking into consideration that the physics concept has not changed.

Imaging and analysis Process. Available at: https://britastro.org/section_information_/exoplanets-section-overview/exoplanet-transit-imaging-and-analysis-process (Accessed: 19 January 2024).	**Relevance:** Very relevant to my project because it offers an illustration of a transit cycle, which I was describing. **Authority:** The article was published by the British Astronomical Association, which is a reputable source in the aspect of astronomy. **Accuracy:** Very accurate because written by people who are experts in the field and members of the association.
American Museum of Natural History (2024). *Guided Exploration: Stars Zone.* Available at: https://www.amnh.org/exhibitions/permanent/the-universe/educator-resources/guided-exploration-stars-zone (Accessed: 19 January 2024).	**Currency:** It is the information presented nowadays in the museum; thus, the source is current. **Relevance:** As I was explaining the marginal dip in the luminosity of a star when a planet blocks the centre of the star, this source became relevant because it explains that all of the energy is generated at the centre. **Authority:** The American Museum of Natural History is a famous museum which indicates that the information presented by it is valid and the source can be considered reputable. **Accuracy:** The museum presents this information to the visitors; therefore, it can be considered reliable and accurate.
European Space Agency (2019). *How to find an exoplanet.* Available at: https://www.esa.int/Science_Exploration/Space_Science/Exoplanets/How_to_find_an_exoplanet (Accessed: 22 December 2023).	**Currency:** 2022 – quite recent, taking into consideration that the physics concept has not changed and only slight advancements have occurred. **Relevance:** Directly relevant to my essay due to the fact that I was explaining why I chose to focus on only one method of exoplanet detection, providing information about the other techniques. **Authority:** The article was published by the European Space Agency, which is a reputable and official source in the aspect of astronomy and space exploration. **Accuracy:** Very accurate as written by people who are experts in the field. Everything is based on facts and is strictly scientifically proven, making it a very reliable source.
Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep Learning*. s.l.:MIT Press.	**Currency:** 2016 – slightly outdated because there may be some new technologies as the field of ML advances, especially in the DNN aspect. Nevertheless, the information taken was fundamental and has not changed. **Relevance:** Extremely relevant and suitable for my dissertation because of such topics as ML experience types and DNNs. The book perfectly addresses my needs. **Authority:** The main author is Ian Goodfellow, who works at Google DeepMind, a leader in AI. He worked on 186 research papers, acquiring a remarkable h-index of 91. Other authors also possess strong expertise in the subject. **Accuracy:** Very accurate. Everything is based on facts and is strictly scientifically proven, making it a very reliable source. It was published by MIT Press indicating a high standard of the book.

Gotame, R. C. (2020). *Transit Photometry Method for Finding the Exoplanets*. Available at: https://physicsfeed.com/post/transit-photometry-method-finding-exoplanets/ (Accessed: 23 December 2023).	**Currency:** 2020 – very acceptable because the fundamental principles stay constant. **Relevance:** Quite relevant for me because I was discussing the benefits and drawbacks of the transit photometry method. **Authority:** The author holds a MS in physics, and is an expert in the field of astrophysics. He launched a physics website to blog where he discusses current advancements. **Accuracy:** No crucial inaccuracies. The article is high-quality and explains the topic in detail.
Hippke, M. and Heller, R. (2019). 'Optimized transit detection algorithm to search for periodic transits of small planets', *Astronomy&Astrophysics*, 623, pp.1-13. doi: https://doi.org/10.1051/0004-6361/201834672	**Currency:** 2019 – good. This article does not need to be very recent because it proposes a new mathematical model for the exoplanet search, so it is not connected to ML. Nevertheless, it makes a reference to a 2018 paper about a new ML model. **Relevance:** Directly related to my dissertation as the research focuses on a newly developed model for exoplanet detection. I studied this paper to obtain an awareness of the mathematical models existing for this task. **Authority:** Both researchers are scientists working in Germany. Michael Hippke is based at Sonneberg Observatory, while René Heller is based at the Max Planck Institute for Solar System Research. Both scientists are experts in the field of astronomy. **Accuracy:** Everything is based on facts and is strictly scientifically proven, making it a very reliable and accurate source. Additionally, the paper was published by a reputable journal 'Astronomy&Astrophysics' meaning a potential peer review process was done, ensuring accuracy.
IBM (2023). *What is machine learning?*. Available at: https://www.ibm.com/topics/machine-learning (Accessed: 5 January 2024).	**Currency:** As the company is large, fundamentally important articles such as 'What is ML' are checked and updated regularly, presenting recent content. **Relevance:** The web page was relevant to me because it supported my point of saying that in this context, AI and ML can be used interchangeably as ML is a huge branch of AI. **Authority:** IBM is a leading and very reputable technological company, so the information was taken from an official and reliable source. **Accuracy:** Correlating with the currency, the accuracy of such pages is perfect as the company is large and possesses great expertise.
Balter, A. (2022). *What is Astrophysics?*. Available at: https://www.space.com/26218-astrophysics.html (Accessed: 7 December 2023).	**Currency:** 2022 – quite recent, and acceptable in the context of astrophysics. **Relevance:** Not directly related to my essay, but was useful for the purpose of introducing astrophysics to the reader. **Authority:** The article was published on the reputable website 'space.com' which focuses primarily on space and astronomy. As astrophysics is under the umbrella of astronomy, the source can be considered trustworthy. **Accuracy:** The article is written with proper referencing to other articles, accurately presenting the content.

Jung, A. (2022). *Machine Learning: The Basics*. Singapore: Springer.	**Currency:** 2022 – good because the ML field advances rather quickly, but it should not have affected this essay because it does not cover cutting-edge topics. **Relevance:** Very relevant to my dissertation because I needed to explore the different types of ML experiences, and touch upon NNs. **Authority:** Alexander Jung received MS and PhD degrees in electrical engineering and signal processing from the Technical University of Vienna. He is a tenured Associate Professor of Machine Learning and leads the 'Machine Learning for Big Data' group at Aalto University's Department of Computer Science. He is an expert in the field with many years of experience. **Accuracy:** Everything is based on facts and is strictly scientifically proven, making it a very reliable and accurate source.
Kapitanova, K., Son, S., Hu, F. (ed) and Hao, Q. (ed) (2012). *Intelligent Sensor Networks*. New York: CRC Press.	**Currency:** 2012 – quite old, but as the basics of ML have not changed, it is acceptable. **Relevance:** Not directly relevant. The book is about sensor networks and signal processing, but the first chapter is dedicated to covering the Machine Learning basics, which is suitable for my objectives. **Authority:** Both Krasimira Kapitanova and Sang Son are scientists affiliated with the University of Virginia with the experience of publishing scholarly works and are members of the highly recognised Institute of Electrical and Electronics Engineers. **Accuracy:** No crucial inaccuracies. Everything is based on facts and is strictly scientifically proven, making it a very reliable and accurate source.
Knight, W. (2017). *The Dark Secret at the Heart of AI*. Available at: https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/ (Accessed: 6 February 2024).	**Currency:** 2017 – quite old for an article about NNs, but as the taken information remains in the same state, it is acceptable. **Relevance:** Not directly relevant to the whole project, but very relevant to my point of stating that NNs are difficult to look into and control. **Authority:** Will Knight is a Senior Editor for AI of a very reputable website 'MIT Technological Review', which is closely connected to the university. He has written over 600 posts, which indicates that he is truly knowledgeable in this sphere. **Accuracy:** The article is high-quality and explains the topic in detail. Everything is based on facts, making it a reliable and accurate source.
Kumar, H. (2018). *How deep should neural nets be?*. Available at: https://kharshit.github.io/blog/2018/04/27/how-deep-should-neual-nets-be (Accessed: 28 January 2024).	**Currency:** 2018 – still acceptable because the structure of the NNs has not changed. **Relevance:** It was related to some extent as I took the picture of the comparison of the layer structure of NNs and DNNs. **Authority:** Harshit Kumar holds a bachelor degree in Computer Science and a MA in Artificial Intelligence. He is an experienced programmer with several years of experience, which demonstrates his expertise in the topic of ML. **Accuracy:** Everything is based on facts with proper referencing, making it a reliable and accurate source.
Lissauer, J. J. (2024). *Extrasolar planet*. Available at:	**Currency:** 2024 – perfect because it is completely recent. **Relevance:** Quite relevant to my project as it provides basic information about exoplanets, which were introduced in the beginning.

https://www.britannica.com/science/extrasolar-planet (Accessed: 17 December 2023).	**Authority:** The writer of the article is a space researcher affiliated with NASA Ames Research Centre. The article was fast-checked by the editors of encyclopaedia Britannica, who are very knowledgeable in the field. Moreover, the website 'Britannica' is itself a very reputable source. **Accuracy:** No crucial inaccuracies. Everything is based on facts with proper referencing, making it a very reliable and accurate source.
Mislis, D., Pyrzas, S. and Alsubai, A. (2018). 'TSARDI: a Machine Learning data rejection algorithm for transiting exoplanet light curves'. *Monthly Notices of the Royal Astronomical Society*, 481(2), pp.1624-1630. doi: https://doi.org/10.1093/mnras/sty2361	**Currency:** 2018 – the researchers were testing the ML models in the application to exoplanet search on a clustering algorithm that has not progressed much, so it is acceptable. **Relevance:** Directly related to my dissertation as the research focuses on a newly developed model for exoplanet detection. My aim was exactly to analyse various studies of ML's application to gain an understanding of the ways it can be done. **Authority:** All three scientists are affiliated with Qatar Environment and Energy Research Institute. They have around 50 or above scholarly works done, with quite high h-indices. **Accuracy:** Everything in the scientific paper is based on facts with proper referencing done, making it a very reliable and accurate source. Additionally, the paper was published by a reputable journal 'Monthly Notices of the Royal Astronomical Society' meaning a potential peer review process was done, ensuring the accuracy.
MIT and the Centre for Astrophysics (2018). *TESS Mission.* Available at: https://tess.mit.edu/science/ (Accessed: 18 December 2023).	**Currency:** 2018 – not very recent but it is not necessary because the website is designed to describe the TESS mission spacecraft which was launched in 2018. **Relevance:** Not directly relevant to my dissertation, but was quite useful for the part about how stars are observed by space observatories. **Authority and Accuracy:** It is the official website of the MIT-led NASA mission meaning the information is very authentic, reliable, and accurate.
Mitchell, T. M. (1997). *Machine Learning.* New York: McGraw-Hill.	**Currency:** 1997 – very old, however, it is still acceptable because only the definition of ML was taken from it. **Relevance:** Loosely relevant as only a high-quality definition of ML was required for the essay. **Authority:** Tom Mitchell is an extremely successful and prominent computer scientist. He has done numerous scholarly works, achieving an outstanding h-index of 100. He is a founder and a former Chair of the Machine Learning Department at a famous Carnegie Mellon University, which highlights his expertise and importance. **Accuracy:** Goodfellow et al. (2016), which is a reputable source, utilised the definition of ML offered by Mitchell, T. M. (1997) which indicates that the stated information remained its quality and is still accurate.
NASA Exoplanet Exploration (2024). *5 Ways to Find a Planet.* Available at: https://exoplanets.nasa.	**Currency:** 2024 – completely recent, and it is updated regularly as it is one of the most essential web pages. **Relevance:** Very related and helpful for my dissertation because I carefully investigated the principle of operation of the transit photometry with emerging problems, as well as the other four techniques.

gov/alien-worlds/ways-to-find-a-planet/ (Accessed: 18 December 2023).	**Authority:** Great - a website of a leading space agency NASA, which is renowned for its discoveries and missions. **Accuracy:** The information stated is regularly updated by experts and provides cutting-edge news making the source extremely reliable and accurate.
Pearson, K. A. (2019). 'A Search for Multiplanet Systems with TESS Using a Bayesian N-body Retrieval and Machine Learning'. *The Astronomical Journal*, 158(6), pp.243-260. doi: https://doi.org/10.3847/1538-3881/ab4e1c	**Currency:** 2019 – some important changes might have happened in the field, but it is still acceptable. **Relevance:** Directly related to my dissertation as the research focuses on a newly developed model for exoplanet detection. My aim was exactly to analyse various studies of ML's application to gain an understanding of the ways it can be done. **Authority:** Kyle Pearson is affiliated with the Lunar and Planetary Laboratory, University of Arizona. He is an expert in the field of astronomy with several scientific papers written. **Accuracy:** Everything in the scientific paper is based on facts with proper referencing done, making it a reliable and accurate source. Additionally, the paper was published by a reputable 'The Astronomical Journal' meaning a potential peer review process was done, ensuring the accuracy.
Pearson, K. A., Palafox, L. and Griffith, C. A. (2018). 'Searching for exoplanets using artificial intelligence'. *Monthly Notices of the Royal Astronomical Society*, 474(1), pp.478-491. doi: https://doi.org/10.1093/mnras/stx2761	**Currency:** 2018– quite old but acceptable because it is one of the first works done on this topic. **Relevance:** Directly related to my dissertation as the research focuses on a newly developed model for exoplanet detection. My aim was exactly to analyse various studies of ML's application to gain an understanding of the ways it can be done. **Authority:** The academic paper was done by Kyle Pearson (above) and two other scientists also affiliated with the Lunar and Planetary Laboratory, University of Arizona, indicating a high level of expertise present. **Accuracy:** Everything in the scientific paper is based on facts with proper referencing done, making it a very reliable and accurate source. Additionally, the paper was published by a reputable journal 'Monthly Notices of the Royal Astronomical Society' meaning a potential peer review process was done, ensuring the accuracy.
Santerne, A. et al. (2012). 'SOPHIE velocimetry of Kepler transit candidates'. *Astronomy&Astrophysics*, 545, pp.1-16. doi: https://doi.org/10.1051/0004-6361/201219608	**Currency:** 2012 – quite old, so the data might be outdated now because a huge advancement in accuracy happened. **Relevance:** Not directly relevant, but is helpful in the context of saying that a further investigation of a predicted exoplanet has to be done because of a potentially high false-positive rate. **Authority:** Good - research was done by several scientists from leading universities in France and Portugal. **Accuracy:** Everything in the scientific paper is based on facts with proper referencing done, making it a reliable and accurate source to that date. Additionally, the paper was published by a reputable journal 'Astronomy&Astrophysics' meaning a potential peer review process was done, ensuring the accuracy.
Seager, S. (2015). 'Searching for Other	**Currency:** 2015 – not very current but acceptable because the further investigation of exoplanets remains the same.

Earths'. *The New Atlantis*, 47, pp.67-75. https://www.jstor.org/stable/43671541	**Relevance:** Loosely relevant as I discussed the further analysis of the planet after it was discovered. **Authority:** Great – Sara Seager is an MIT professor with many years of experience in the field of astronomy and exoplanets specifically, having a BSc and a PhD in mathematics & physics and astronomy respectively. She wrote several books and made exoplanet discoveries, as well as formed a Seager equation. **Accuracy:** Everything is accurately and carefully described and explained.
Tey, E. et al. (2023). 'Identifying Exoplanets with Deep Learning. V. Improved Light-curve Classification for TESS Full-frame Image Observations'. *The Astronomical Journal*, 165(3), pp. 165-195. doi: https://doi.org/10.3847/1538-3881/acad85	**Currency:** 2023 – very recent and applicable to today's world. **Relevance:** Directly related to my dissertation as the research focuses on a newly developed model for exoplanet detection. My aim was exactly to analyse various studies of ML's application to gain an understanding of the ways it can be done. **Authority:** Great – many famous scientists worked on this paper that are connected to the best American universities such as MIT and Princeton that perform cutting-edge research, making the paper trustworthy. Sara Seager (above source) is one of the people who contributed to the paper. **Accuracy:** Everything in the scientific paper is based on facts with proper referencing done, making it a very reliable and accurate source. Additionally, the paper was published by a reputable 'The Astronomical Journal' meaning a potential peer review process was done, ensuring the accuracy.
The International Astronomical Union (2018). *Pluto and the Developing Landscape of Our Solar System.* Available at: https://www.iau.org/public/themes/pluto/ (Accessed: 17 December 2023).	**Currency:** 2018 – good because the definition of a planet does not need to be changed once defined. **Relevance:** Not directly relevant to my essay, however, a strict definition of a planet was necessary for a better understanding. **Authority:** Great - The International Astronomical Union was founded in 1919 and consists of over 12,000 professional astronomers. The definition of a planet was set in 2006 at the Prague General Assembly, having a vote across the union's members. **Accuracy:** The information stated in the source is completely accurate as it was written and verified by professionals.
The Planetary Society (2020). *Down in Front!: The Transit Photometry Method.* Available at: https://www.planetary.org/articles/down-in-front-the-transit-photometry-method (Accessed: 22 December 2023).	**Currency:** 2020 – good. It introduces the concept of transit photometry, which has not changed. **Relevance:** Very relevant as I needed to talk about some specific details about the method such as the effectiveness, multiwavelength observations, and timings of staring at stars. **Authority:** Great - The planetary society was founded in the early 1980s by three well-known scientists Bruce Murray, Carl Sagan, and Louis Friedman. The website is run by experts in the field of exoplanets. **Accuracy:** No obvious inaccuracies. The article is high-quality and explains the topic in detail. Additionally, it was checked by several professional members of the society.

The Planetary Society, n.d. *Exoplanets, worlds orbiting other stars.* Available at: https://www.planetary.org/worlds/exoplanets (Accessed: 17 December 2023).	**Currency:** Although the article has no gate, it states that by the date of writing the article, 5000 exoplanets were discovered, which means that it is certainly more recent than 2020 making it reasonably current. **Relevance:** Quite relevant, but the only information required was about the fact that discovering other star systems allows us to better understand the formation of the universe. **Authority:** Great - The planetary society was founded in the early 1980s by three well-known scientists Bruce Murray, Carl Sagan, and Louis Friedman. The website is run by experts in the field of exoplanets. **Accuracy:** No crucial inaccuracies. The article is high-quality and explains the topic in detail.
Wang, K., Ge, J., Willis, K. and Zhao, Y. (2024). 'The GPU phase folding and deep learning method for detecting exoplanet transits. *Monthly Notices of the Royal Astronomical Society*, 528(3), pp.4053-4067. doi: https://doi.org/10.1093/mnras/stae245	**Currency:** 2024 – perfect. It is cutting-edge research. **Relevance:** Directly related to my dissertation as the research focuses on a newly developed model for exoplanet detection. My aim was exactly to analyse various studies of ML's application to gain an understanding of the ways it can be done. **Authority:** Great – experienced researchers who are all affiliated with leading universities across the world. They are professionals in the field. **Accuracy:** No obvious inaccuracies. Everything in the scientific paper is based on facts with proper referencing done, making it a very reliable and accurate source. Additionally, the paper was published by a reputable journal 'Monthly Notices of the Royal Astronomical Society' meaning a potential peer review process was done, ensuring the accuracy.
Webb Space Telescope (2024). *About the Telescope*. Available at: https://webbtelescope.org/about (Accessed: 18 December 2023).	**Currency:** 2024 – very recent, and it is updated regularly as Webb is a current mission aimed at various discoveries. **Relevance:** Very related and helpful for my dissertation because I carefully investigated the principle of operation of the transit photometry with emerging problems, as well as the other four techniques. **Authority and Accuracy:** It is the official website of the NASA mission meaning the information is very authentic, reliable, and accurate. The information stated is regularly updated by experts and provides cutting-edge news making the source extremely reliable and accurate.
Wells, R. (2020). *The properties of transiting exoplanets*. Belfast: Queen's University Belfast.	**Currency:** 2020 – good and reasonably recent. For the purpose of comparing the two maths models, it is perfect. **Relevance:** Directly relevant to my project, but I only utilised a small part of this paper for the purpose of showing the difference between the principles of operation of BLS and TLS models. **Authority:** Good - Robert Wells is a professor at the well-known Queen's University Belfast. He is an expert in the sphere of astronomy. **Accuracy:** No obvious inaccuracies. Everything in the scientific paper is based on facts with proper referencing done, making it a very reliable and accurate source. It was published by Queen's University Belfast indicating a high standard of the book.

Wenz, J. (2023). *How the first exoplanets were discovered.* Available at: https://www.astronomy.com/science/how-the-first-exoplanets-were-discovered/ (Accessed: 18 December 2023).	**Currency:** Great because the information taken is historical and does not change. **Relevance:** Not directly relevant to my dissertation, but it was useful for providing background information about the first exoplanet discovered. **Authority:** Great - the article was published on the reputable website 'astronomy.com', which publishes the world's best-selling astronomy magazine. The writer is a freelance science writer with particular expertise in astronomy. **Accuracy:** No crucial inaccuracies. The article is high-quality and explains the topic in detail.
Williams, M. (2017). *What is the Transit Method?*. Available at: https://www.universetoday.com/137480/what-is-the-transit-method/ (Accessed: 23 December 2023).	**Currency:** 2017 – a little bit outdated, but it does not affect this dissertation because the required information has not changed. **Relevance:** Quite relevant for me because I was discussing the advantages and disadvantages of the transit photometry method. **Authority:** Good - the article was published on a website specialising in astronomy, and the author has expertise in the subject. **Accuracy:** No obvious inaccuracies. The article is high-quality and explains the topic in detail, properly referencing utilised sources.
Winn, J. (2023). *The Little Book of Exoplanets.* s.l.:Princeton University Press.	**Currency:** 2023 – great. Contains recent discoveries and news regarding exoplanet identification. **Relevance:** Directly relevant to my project, but I only utilised a small part of this book for the purpose of demonstrating that modern space missions are beneficial in terms of exoplanet detection. **Authority:** Joshua Winn is a professor of astrophysical sciences at Princeton University. He is also a member of the TESS mission. He discovered several exoplanets with numerous scholarly works done, achieving a remarkable h-index of 99. **Accuracy:** Everything in the book is based on facts with proper referencing done, making it an extremely reliable and accurate source. It was published by Princeton University Press indicating a high standard of the book.
Yeh, L.-C. and Jiang, I.-G. (2021). 'Searching for Possible Exoplanet Transits from BRITE Data through a Machine Learning Technique'. *The Astronomical Society of the Pacific*, 133, pp.1-12. doi: https://doi.org/10.1088/1538-3873/abbb24	**Currency:** 2021 – decent and completely acceptable. **Relevance:** Directly related to my dissertation as the research focuses on a newly developed model for exoplanet detection. My aim was exactly to analyse various studies of ML's application to gain an understanding of the ways it can be done. **Authority:** Good - both researchers are affiliated with National Tsing Hua University. They are experts in the fields of astronomy and astrophysics. **Accuracy:** Everything in the book is based on facts with proper referencing done, making it a very reliable and accurate source. Additionally, the paper was published by a reputable journal 'The Astronomical Society of the Pacific' meaning a potential peer review process was done, ensuring the accuracy.

Zucker, S. and Giryes, R. (2018). 'Shallow Transits—Deep Learning. I. Feasibility Study of Deep Learning to Detect'. *The American Astronomical Society*, 155, pp.147-155. doi: https://doi.org/10.3847/1538-3881/aaae05	**Currency:** 2018 – a little bit outdated because DNNs might have advanced since then. **Relevance:** Directly related to my dissertation as the research focuses on a newly developed model for exoplanet detection. My aim was exactly to analyse various studies of ML's application to gain an understanding of the ways it can be done. **Authority:** Good – the authors are very knowledgeable scientists with many years of experience affiliated with Tel Aviv University. They have written numerous academic papers achieving very high h-indices. **Accuracy:** Good – everything in the book is based on facts with proper referencing done, making it a very reliable and accurate source. Additionally, the paper was published by a reputable journal 'The American Astronomical Society' meaning a potential peer review process was done, ensuring the accuracy.